

1-2008

## Why Binomial Distributions Do Not Work as Proof of Employment Discrimination

Ben Ikuta

Follow this and additional works at: [https://repository.uchastings.edu/hastings\\_law\\_journal](https://repository.uchastings.edu/hastings_law_journal)



Part of the [Law Commons](#)

---

### Recommended Citation

Ben Ikuta, *Why Binomial Distributions Do Not Work as Proof of Employment Discrimination*, 59 HASTINGS L.J. 1235 (2008).  
Available at: [https://repository.uchastings.edu/hastings\\_law\\_journal/vol59/iss5/10](https://repository.uchastings.edu/hastings_law_journal/vol59/iss5/10)

This Note is brought to you for free and open access by the Law Journals at UC Hastings Scholarship Repository. It has been accepted for inclusion in Hastings Law Journal by an authorized editor of UC Hastings Scholarship Repository. For more information, please contact [wangangela@uchastings.edu](mailto:wangangela@uchastings.edu).

# Why Binomial Distributions Do Not Work as Proof of Employment Discrimination

BEN IKUTA\*

## INTRODUCTION

In employment discrimination cases under Title VII,<sup>1</sup> “[n]o issue has been more stubborn than deciding to what extent a plaintiff must present specific evidence about a defendant’s state of mind.”<sup>2</sup> Many times, an employer’s violation of Title VII affects the discriminated class as a whole, not just the protected individual.<sup>3</sup> The plaintiff in these “systemic disparate treatment” cases must prove the elements of a *prima facie* case, which requires that he or she presents a set of facts that raise “an inference of discrimination . . . because these acts, if otherwise unexplained, are more likely than not based on the consideration of impermissible factors.”<sup>4</sup> However, contrary to individual disparate treatment cases,<sup>5</sup> specific and isolated proof of discrimination is usually not sufficient in systemic disparate treatment cases to show a *prima facie* case since discrimination must be “the [defendant’s] standard operating procedure—the regular rather than the unusual practice.”<sup>6</sup> Additionally,

---

\* J.D., University of California, Hastings College of the Law, 2008. I would like to thank Jennifer Luczkowiak, Pilar Stillwater, Erik Christensen, Jay Nelson, Katie Annand, Scott Dommies, and Amber Jones for their invaluable assistance on this Note.

1. Title VII prohibits discrimination by employers on the basis of race, color, religion, sex, or national origin. See 42 U.S.C. § 2000e (2000). There are also other ways to unlawfully discriminate. E.g., The Age Discrimination in Employment Act, 29 U.S.C. §§ 621–34.

2. ROBERT N. COVINGTON & KURT H. DECKER, *INDIVIDUAL EMPLOYEE RIGHTS* 159 (1995).

3. THOMAS R. HAGGARD, *UNDERSTANDING EMPLOYMENT DISCRIMINATION* 83 (2001).

4. *Furnco Constr. Corp. v. Waters*, 438 U.S. 567, 577 (1978). This Note will focus on statistical analysis in systemic disparate treatment cases, not systemic disparate impact cases. See Elaine W. Shoben, *Differential Pass-Fail Rates in Employment Testing: Statistical Proof Under Title VII*, 91 HARV. L. REV. 793 (1978), for a good description of statistical analysis in systemic disparate impact cases.

5. Proof of a *prima facie* case in individual disparate treatment cases involves four distinct factors:

(i) that [the applicant] belongs to a [ ] minority; (ii) that he applied and was qualified for a job for which the employer was seeking applicants; (iii) that, despite his qualifications, he was rejected; and (iv) that, after his rejection, the position remained open and the employer continued to seek applicants from persons of complainant’s qualifications.

*McDonnell Douglas v. Green*, 411 U.S. 792, 802 (1973).

6. *Int’l Bhd. of Teamsters v. United States*, 431 U.S. 324, 336 (1977); see also RAMONA L.

not only are individual acts of discrimination insufficient, but “[a]necdotal evidence of individual class members is not necessary.”<sup>7</sup> In these cases, an illustration of a general trend toward discrimination is needed to establish a *prima facie* case while individual evidence of discrimination is used only as secondary or persuasive weight, if at all.<sup>8</sup>

Although it is possible for employer discrimination policies to be facially apparent,<sup>9</sup> usually employer policies are not overtly discriminatory.<sup>10</sup> In those situations where it is not apparent, the only method of proving a *prima facie* case of discrimination in most systemic disparate treatment instances is to show a pattern of discrimination in the employer’s hiring decisions.<sup>11</sup> Since these patterns can usually “be identified and analyzed in quantitative form,”<sup>12</sup> cases rely primarily, and often exclusively, on statistics.<sup>13</sup> If the plaintiff offers statistics that indicate a hiring practice that is unlikely absent discrimination, then the plaintiff is deemed to have created a *prima facie* case, thereby shifting to the defendant the burden of rebutting the plaintiff’s statistical proof.<sup>14</sup> The defendant can rebut this presumption by: (1) demonstrating that the plaintiff’s statistics are “inaccurate or insignificant”;<sup>15</sup> (2) offering his own statistical proof of nondiscriminatory hiring practices;<sup>16</sup> or (3) offering

PAETZOLD & STEVEN L. WILLBORN, *THE STATISTICS OF DISCRIMINATION: USING STATISTICAL EVIDENCE IN DISCRIMINATION CASES* 1 (1994) (illustrating that individual acts of discrimination “[are] not enough to meet this standard”).

7. ROBERT BELTON, *REMEDIES IN EMPLOYMENT DISCRIMINATION LAW* 49 n.89 (1992); *accord* PAETZOLD & WILLBORN, *supra* note 6; MARK A. ROTHSTEIN & LANCE LIEBMAN, *EMPLOYMENT LAW: CASES AND MATERIALS* 154 (6th ed. 2007). “Some cases have indicated that statistics alone can create a *prima facie* case of intentional group exclusion.” *Id.*; *see also, e.g.*, *EEOC v. Am. Nat’l Bank*, 652 F.2d 1176, 1190 (4th Cir. 1981); *United States v. Sheet Metal Workers Int’l Ass’n Local 36*, 416 F.2d 123, 127 n.7 (8th Cir. 1969).

8. BELTON, *supra* note 7, at 49.

9. *E.g.*, *L.A. Dep’t of Water & Power v. Manhart*, 435 U.S. 702, 702 (1978) (requiring female employees to make larger contributions to the pension fund than male employees).

10. RICHARD R. CARLSON, *EMPLOYMENT LAW* 114 (2005) (“[D]iscrimination frequently takes subtle forms. An employer who knows discrimination is illegal is not likely to announce his bias as a matter of corporate policy. Moreover, bias might affect one supervisor or manager but not others, so that the effect of the bias is not reflected in overall employment statistics for the entire company.”); MACK A. PLAYER, *EMPLOYMENT DISCRIMINATION LAW* 343 (1988) (“The plaintiff may find ‘smoking gun’ evidence, such as a written memorandum describing the [discriminatory] policy, but this is not likely.”).

11. *See* HAGGARD, *supra* note 3; PAETZOLD & WILLBORN, *supra* note 6; Louis J. Braun, *Statistics and the Law: Hypothesis Testing and Its Application to Title VII Cases*, 32 HASTINGS L.J. 59, 59–61 (1980). In this Note, the word “hiring” will be used for the purpose of simplicity. However, the same analysis could and should be used for others instances of employment discrimination.

12. PAETZOLD & WILLBORN, *supra* note 6.

13. *E.g.*, *United States v. Sheet Metal Workers Int’l Ass’n Local 36*, 416 F.2d 123 (8th Cir. 1969); *see* HAGGARD, *supra* note 3, at 83–84; PAETZOLD & WILLBORN, *supra* note 6, at 1–2.

14. *See* Int’l Bhd. of Teamsters v. United States, 431 U.S. 324, 339 (1977).

15. *Id.* at 360.

16. HAGGARD, *supra* note 3, at 85.

nondiscriminatory explanations for the disparity in the statistics.<sup>17</sup>

While courts sometimes view the statistical data initially presented as so one-sided that little further analysis is required,<sup>18</sup> courts are more inclined to use complicated statistical methods to determine whether the statistical discrepancy in the underhiring of a protected class is due to pure chance, or rather discriminatory reasons.<sup>19</sup> A majority of courts use standard deviations of normal distribution approximations based on binomial distributions as described in *Hazelwood School District v. United States*.<sup>20</sup> This Note aims to prove that the assumption of independent trials in the usage of binomial distributions is unjustified in employment discrimination cases, and is therefore an inappropriate and inaccurate model to use as a primary basis for systemic disparate treatment cases.

Part I of this Note will briefly describe the background of *Teamsters* and *Hazelwood*, the two landmark cases dealing with statistics in systemic employment cases, and will also explain the statistical analysis used in the jury selection case, *Castaneda v. Partida*,<sup>21</sup> upon which *Teamsters* and *Hazelwood* relied heavily. Part II will thoroughly explain the binomial theorem and the reasons for its use within the systemic employment law context. Part III will describe why employment decisions are not independent trials for employers, and why other evidence should have more influence, as binomial distributions are not an accurate tool in showing discrimination.

## I. THE EFFECT OF THE *TEAMSTERS* AND *HAZELWOOD* CASES ON THE USE OF STATISTICS IN SYSTEMIC DISPARATE TREATMENT CASES

*Teamsters* and *Hazelwood* created the backbone for using statistical data to establish a *prima facie* case of discrimination in systemic disparate

---

17. EEOC v. Chicago Miniature Lamp Works, 947 F.2d 292, 302 (7th Cir. 1991) (explaining that a low percentage of African-American employees was due to the fact that the job did not require fluency in English, which made the job more attractive to non-English speaking Asians and Hispanics who applied in disproportionate numbers). However, explaining away the statistics by reference to the inability of minorities to pass a certain test might open the employer to disparate impact discrimination. See HAGGARD, *supra* note 3, at 84.

18. See, e.g., *Teamsters*, 431 U.S. at 336-41; EEOC v. Am. Nat'l Bank, 652 F.2d 1176, 1190 (4th Cir. 1981) (focusing on the underrepresentation of African-Americans in office, clerical, and management jobs); see also Thomas J. Campbell, *Regression Analysis in Title VII Cases: Minimum Standards, Comparable Worth, and Other Issues Where Law and Statistics Meet*, 36 STAN. L. REV. 1299, 1299 (1984).

19. See, e.g., *Wilkins v. Univ. of Houston*, 654 F.2d 388 (5th Cir. 1981); *Am. Nat'l Bank*, 652 F.2d at 1176; *Hameed v. Int'l Ass'n of Bridge, Structural & Iron Workers Local 396*, 637 F.2d 506 (8th Cir. 1980); *Bd. of Educ. v. Califano*, 584 F.2d 576 (2d Cir. 1978); *Otero v. Mesa County Valley Sch. Dist.*, 568 F.2d 1312 (10th Cir. 1977).

20. 433 U.S. 299 (1977); see also *Wilkins*, 654 F.2d at 397; *Am. Nat'l Bank*, 652 F.2d at 1176; *Hameed*, 637 F.2d at 514; *Califano*, 584 F.2d at 585; *Otero*, 568 F.2d at 1316.

21. 430 U.S. 482 (1977).

treatment cases<sup>22</sup> and further chose the appropriate statistics and comparisons for determining the existence of employment discrimination. The Court in both *Teamsters* and *Hazelwood* justified its use of statistics by reference to *Castaneda v. Partida*, a jury selection case in which the Court endorsed the use of statistics to determine discrimination.<sup>23</sup>

A. *INTERNATIONAL BROTHERHOOD OF TEAMSTERS V. UNITED STATES*

In *Teamsters*, the central claim was that the employer had engaged in a pattern or practice of discriminating against minorities in "line driving" positions.<sup>24</sup> The company had three main employee positions: "line drivers," "servicemen," and "city operations."<sup>25</sup> Although none of these positions required a formal education, "line drivers" were paid a higher salary than the other positions.<sup>26</sup> In analyzing the statistical data, the Court observed that of 571 minority employees, less than 3% held "line driving" positions and only 39% of the nonminority employees held the two lower paying positions.<sup>27</sup> Rejecting the defendant's argument that "statistics can never in and of themselves prove the existence of a pattern or practice of discrimination, or even establish a prima facie case," the Court held that discrimination should be inferred "where it reached proportions comparable to those in this case."<sup>28</sup>

However, the most widely followed section is the analysis in footnote seventeen, which compares the ratio of minorities to nonminorities in the "line driver" positions to the ratio of minorities to nonminorities in the general population of the cities where each of the terminals were located.<sup>29</sup> The Court reasoned that if 17.88% of Los Angeles is African-American, then the percentage of African-Americans in a low-education job like "line driving" should be comparable to

---

22. See HAGGARD, *supra* note 3, at 83-84; PAETZOLD & WILLBORN, *supra* note 6, at 1-2; Thomas J. Sugrue & William B. Fairley, *A Case of Unexamined Assumptions: The Use and Misuse of the Statistical Analysis of Castaneda/Hazelwood in Discrimination Litigation*, 24 B.C. L. REV. 925, 925-27 (1983).

23. See BELTON, *supra* note 7, at 226 ("Courts often cite *Castaneda* to justify the use of standard deviation analysis in employment discrimination cases.").

24. Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 328-29 (1977).

25. *Id.* at 330 n.3.

26. *Id.*

27. *Id.* at 337-38. Nationwide, the employer had 1,802 line drivers, all of whom were white except for thirteen minority drivers. In the less desirable city driving jobs, there were 1,117 white employees and 167 minority employees. *See id.* at 337.

28. *Id.* at 339; *see also id.* at 340 n.20 ("Statistics showing racial or ethnic imbalance are probative in a case such as this one only because such imbalance is often a telltale sign of purposeful discrimination; absent explanation, it is ordinarily to be expected that nondiscriminatory hiring practices will in time result in a work force more or less representative of the racial and ethnic composition of the population in the community from which employees are hired.").

29. *Id.* at 338 n.17.

17.88%.<sup>30</sup> Although these comparisons were extremely significant in illustrating what the appropriate statistical comparisons should be, the Court did not address the degree of statistical discrepancy needed to find discrimination because the statistical evidence was so one-sided.<sup>31</sup>

B. *HAZELWOOD SCHOOL DISTRICT V. UNITED STATES*

The problem left open in *Teamsters* concerning the amount of statistical discrepancy needed to raise an inference of discrimination was “solved” in *Hazelwood*.<sup>32</sup> *Hazelwood* involved the hiring of minorities for teaching positions in a public school. Similarly to *Teamsters*, the Court in *Hazelwood* used the percentage of minorities in teaching positions in the surrounding areas as a basis of comparison.<sup>33</sup> The Court recognized that the percentage of minority teachers in the surrounding areas could either be 5.7% or 15.4%, depending on whether the court thought that it was appropriate to include one surrounding school district that had chosen to pursue a goal of hiring 50% minority teachers.<sup>34</sup> Observing that the school at issue in the case had only a 3.7% hiring rate for minority teachers, the Court relied on *Castaneda* when it held that using “statistical methodology . . . involving the calculation of the standard deviation as a measure of predicted fluctuations [shows] the difference between using 15.4% and 5.7% as the area-wide figure would be significant.”<sup>35</sup> However, *Hazelwood* did not fully describe the method or the pitfalls of using standard deviation based on binomial distributions.<sup>36</sup>

C. *CASTANEDA V. PARTIDA*

*Castaneda* involved a claim of underrepresentation of Mexican-Americans in grand jury selections in criminal cases.<sup>37</sup> The Court held that because the population of Hidalgo County (over 180,000 people) consisted of 79.1% Mexican-Americans, it would follow logically that close to 79.1% of the 870 people, or 688 people,<sup>38</sup> summoned to serve as

---

30. *Id.*; see also BELTON, *supra* note 7, at 48 (“As the Court stated in *Teamsters*, statistical evidence is often a telltale sign of purposeful discrimination.”).

31. See *Teamsters*, 431 U.S. at 338 n.17; see also MICHAEL ZIMMER ET AL., CASES AND MATERIALS ON EMPLOYMENT DISCRIMINATION 256 (4th ed. 1997) (“*Teamsters* presented a relatively easy statistical case of discrimination. Virtually no minority group members were assigned line-driver positions.”).

32. *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 312 n.17 (1977).

33. *Id.* at 309 n.13. *Hazelwood* explained that *Teamsters* was able to use the general population since the line of employment at issue in that case required very low skill that “one of many persons possess or can fairly readily acquire.” *Id.* However, where “special qualifications are required to fill particular jobs, comparisons to the general population . . . fail to take into account special qualifications for the position in question.” *Id.*; see also Braun, *supra* note 11, at 62–64.

34. *Hazelwood* 433 U.S. at 308–09.

35. *Id.* at 312 n.17; see also *Castaneda v. Partida*, 430 U.S. 482, 496–97 n.17 (1977); ZIMMER ET AL., *supra* note 31, at 289–93.

36. See *Hazelwood*, 433 U.S. at 312 n.17.

37. *Castaneda*, 430 U.S. at 482.

38. This is also known as “expected value.” See MELVIN HAUSNER, ELEMENTARY PROBABILITY

grand jurors over an eleven year period in that same county should also have been Mexican-American.<sup>39</sup> The Court correctly reasoned that some variation from the expected value of 688 would be tolerable since "some fluctuation from the expected number is predicted."<sup>40</sup> The Court also properly recognized that the fluctuation had to be insignificant enough such that "the statistical model shows that the results of a random drawing are likely to fall in the vicinity of the expected value."<sup>41</sup> More concisely, the Court held that the further the actual number of Mexican-Americans deviated from 688 (or 79.1% of the total of 870), the less likely the reason for the variation in the actual number could be attributed to random chance.<sup>42</sup>

*Castaneda* then briefly described how to determine the probability that the actual observed number would deviate from the expected value.<sup>43</sup> The probabilities of drawing different numbers of Mexican-Americans was given by a binomial distribution, and the "measure of the predicted fluctuations from the expected value is the standard deviation, defined for the binomial distribution as the square root of the product of the total number in the sample multiplied by the probability of selecting a Mexican-American multiplied by the probability of selecting a non-Mexican-American."<sup>44</sup> Additionally, the Court found that within this formula, there was a general rule: "[I]f the difference between the expected value and the observed number is greater than two or three standard deviations, then the hypothesis that the . . . drawing was random would be suspect to a social scientist."<sup>45</sup> Two and three standard deviations would indicate a 5% and 0.3% chance, respectively, that the disparity was caused by random sampling and not other reasons such as discrimination.<sup>46</sup> Although these statistics were relatively accurate in *Castaneda* and *Hazelwood*, neither case described the assumptions necessary for the binomial theorem to be an accurate description of the probabilities of discrimination.<sup>47</sup> As a result, lower courts have blindly followed *Castaneda* and *Hazelwood*, and their two to three standard deviation rule in systemic employment discrimination cases by literally

---

THEORY 180 (1971) (illustrating how to derive expected values).

39. *Castaneda*, 430 U.S. at 496 n.17.

40. *Id.*

41. *Id.*

42. *Id.*

43. *Id.*

44. *Id.* For a full description of the binomial method, see *infra* Part II.

45. *Castaneda*, 430 U.S. at 496 n.17; see also *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 312 n.17 (1977); HAUSNER, *supra* note 38, at 244.

46. See *infra* Part II.

47. See PAETZOLD & WILLBORN, *supra* note 6, at 32-33; Paul Meier et al., *What Happened in Hazelwood: Statistics, Employment Discrimination, and the 80% Rule*, 1984 AM. B. FOUND. RES. J. 139; Sugrue & Fairley, *supra* note 22, at 928-29.

plugging numbers into the standard deviation formula without fully comprehending the binomial distribution model and its possible shortcomings.<sup>48</sup>

## II. BINOMIAL DISTRIBUTIONS

In statistics, a binomial distribution is the discrete probability distribution of the number of “successes” in a sequence of independent “yes or no” experiments, each with a definitive probability.<sup>49</sup> The mathematical formula for binomial distributions is described as

$$f(k; n, p) = \binom{n}{k} p^k (1 - p)^{(n-k)}$$

where

$$\binom{n}{k}$$

also known as the “binomial coefficient,” is described as

$$\binom{n}{k} = \frac{n \cdot (n - 1) \cdots (n - k + 1)}{k \cdot (k - 1) \cdots 1} = \frac{n!}{k! (n - k)!}$$

In the formulas above,  $p$  is the probability of a “success,”  $(1 - p)$  is the probability of a “failure,”  $n$  is the total number of yes/no experiments, and  $k$  is the actual number of “successes.”<sup>50</sup>

In addition, there is a critical additional requirement of the binomial theorem that *Hazelwood* and *Castaneda* did not address. All of the trials must be “independent,” meaning that the outcome of any one trial cannot affect any other trial.<sup>51</sup> Although these formulas may be intimidating to less mathematically savvy individuals, they are important for appreciating the shortcomings of using binomial distributions in the employment discrimination context.<sup>52</sup>

48. See Sugrue & Fairley, *supra* note 22, at 926–27; see also, e.g., *Wilkins v. Univ. of Houston*, 654 F.2d 388 (5th Cir. 1981); *EEOC v. Am. Nat'l Bank*, 652 F.2d 1176 (4th Cir. 1981); *Hameed v. Int'l Ass'n of Bridge, Structural & Iron Workers Local 396*, 637 F.2d 506 (8th Cir. 1980); *Bd. of Educ. v. Califano*, 584 F.2d 576 (2d Cir. 1978); *Otero v. Mesa County Valley Sch. Dist.*, 568 F.2d 1312 (10th Cir. 1977). Professor Kaye heavily criticizes using this 5% rule based on a value of two standard deviations and the danger it causes. D.H. Kaye, *Is Proof of Statistical Significance Relevant*, 61 WASH. L. REV. 1333, 1343–44 (1986).

49. HAUSNER, *supra* note 38, at 249. This is simply an introductory probability theory textbook. Almost every introductory probability theory textbook and most introductory statistics textbooks will provide sufficient analysis and support for the mathematical formulas and breakdown in this Note.

50. *Id.* at 61, 249–50.

51. HAUSNER, *supra* note 38, at 249–50; see also PAETZOLD & WILLBORN, *supra* note 6, at 32–33; Meier et al., *supra* note 47, at 157.

52. See Sugrue & Fairley, *supra* note 22, at 926–27.



The best way to illustrate the procedures of the binomial method is by example. Suppose a coin is flipped ten times. Suppose further that if the coin lands heads, it will be deemed a "success," and if the coin lands tails, then it will be deemed a "failure." The expected number of "successes" (heads) would be the probability of heads landing on one particular occasion (50%) multiplied by the total number of experiments (ten), or five heads. Therefore, when the coin ultimately lands with only two heads and eight tails, a suspicion of a biased or "discriminatory" coin might be claimed. In this case, binomial distribution analysis is vital in demonstrating whether the disparity is most likely due to chance or discrimination.

The  $p^k$  element of the binomial distribution formula describes the probability that the first  $k$  trials will result in  $k$  successes.<sup>53</sup> In the coin illustration where there are two successes, this would be the probability that both the first two coin flips would be heads. Since the probability of flipping a heads on any given flip is  $1/2$ , following the product rule of independent trials,<sup>54</sup> the probability that the first two flips of the coin are heads is  $1/2$  multiplied by  $1/2$ , which equals  $1/4$ .<sup>55</sup> Using the formula  $p^k$ , where  $p$  is the probability of a success (50%) and  $k$  is the number of successes (two), the same result of  $1/4$  is obtained.

The  $(1 \cdot p)^{(n-k)}$  section of the binomial distribution formula illustrates the probability that the last  $(n - k)$  trials will result in failures.<sup>56</sup>  $(1 - p)$  represents the probability of a failure on any given trial and  $(n - k)$  symbolizes the total number of failures witnessed. Using our coin example, this is the probability that the last eight flips will be tails. Again using the product rule of independent trials, the probability that the last eight flips will be tails is  $(1/2)^8$ , or  $1/256$ . The formula  $(1 \cdot p)^{(n-k)}$ , where  $p$  is the probability of success (50%),<sup>57</sup>  $n$  is the total number of trials (ten), and  $k$  is the number of successes (two), yields the same result of  $1/256$ . Therefore, the probability of the first two trials being heads and the last eight trials being tails is calculated by multiplying  $1/4$  and  $1/256$ , which equals  $1/1024$ .<sup>58</sup>

Although it is now acknowledged that the probability of ten trials producing two heads and eight tails in that precise order in ten trials is  $1/1024$ , there are various combinations of ten coin flips that will result in

53. HAUSNER, *supra* note 38, at 249-51.

54. *Id.*

55. This can also be shown using common sense. The four equal probabilities of two coin flips are HH, TT, HT, and TH, where H represents heads and T represents tails. Therefore, the probability of HH occurring is  $1/4$ .

56. HAUSNER, *supra* note 38, at 249-51.

57. Notice that  $(1 - p)$  is the equivalent to the probability of a failure on any given trial.

58. Since this is such a simple example, this can also be done using strictly the product rule of independent variables, or 50% to the tenth power. See HAUSNER, *supra* note 38, at 97 (describing the product rule).

two heads and eight tails. For example, the first four flips could be tails, the next two heads, and the last four tails. Therefore, the probability of exactly two heads occurring, regardless of order, in ten coin flips is found by multiplying the probability of one specific occurrence (in this case  $1/1024$ ) by the total possible number of occurrences that would satisfy the final result of two heads. Again this is through use of the product rule.

The total possible number of permutations is described by the

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

component of the binomial distribution theorem, also known as a binomial coefficient. To explain this formula, observe that from the total number of experiments ( $n$ ), the concern is the total number of ways to obtain successes ( $k$ ). Therefore, the first selection is made from the whole set of  $n$  trials. However, once the first selection is made, the second selection is made from  $(n - 1)$  trials, the third selection from  $(n - 2)$  trials, and so on until the last selection is made from the  $(n - (k - 1))$ , or  $(n - k + 1)$ , trials.<sup>59</sup> Therefore, the number of total permutations is described as  $(n \cdot (n - 1) \cdot (n - 2) \dots (n - k + 1))$  or  $(n!/(n - k)!)$ .<sup>60</sup>

This assumes, however, that the objects being chosen have a distinct and particular order. For example, the different ways to choose two letters out of the name "BEN" would be BE, EB, EN, NE, BN, and NB for a total of  $(3!/(3 - 2)!)$ , or six different ways.<sup>61</sup> However, this isn't the appropriate permutation in binomial distribution cases because it does not matter in what order selections are being made.<sup>62</sup> For instance, in systemic employment discrimination cases, if a minority (X) is chosen in the third trial and another minority (Y) is chosen in the seventh trial, it makes no difference if Y is instead chosen third and X chosen seventh.<sup>63</sup> In order to avoid this "double counting," the permutations must be further divided by  $k!$ , the permutations of  $k$  distinct objects.<sup>64</sup> This ultimately gives the formula for the binomial coefficient, or

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

---

59. *Id.* at 250–51.

60. *Id.*

61. *See id.* at 51.

62. *See id.* at 55. Be careful not to get this analysis confused with the independence of trials. The trials are still completely independent.

63. *Id.* This is assuming that the positions are not different. If the positions are different, then binomial distributions would not work.

64. *Id.* at 55–56.

which illustrates the number of ways in which  $k$  successes can be selected from a set of  $n$  independent trials.<sup>65</sup> Therefore, in the BEN example, the only permutations are BE, BN, and EN for a total of three different ways, which is verified by the formula  $(3!/(2! \cdot (3 - 2)!))$ . Likewise, in the coin example above, the total number of ways to obtain two heads from a set of ten independent trials is  $(10!/(2! \cdot 8!))$  or forty-five different ways, which can also be ascertained by simply listing all of the possibilities.<sup>66</sup>

Thus, in the coin example, the probability of drawing *exactly* two heads from a total of ten independent trials is the probability of it happening in a single order, or  $1/1024$ , multiplied by the total number of different ways it is possible, or forty-five, for a probability of  $45/1024$ , or approximately 4.4%, which follows from the basic definition of the binomial distribution. However, to determine if the coin is "discriminatory," the issue is not the probability of drawing *exactly* two heads, but rather the probability of drawing two heads *or fewer*. Therefore, binomial distributions should also be utilized to determine the probability of drawing one head, which in this example is approximately 1.0%,<sup>67</sup> and the probability of drawing no heads at all, which would be around 0.1%.<sup>68</sup> Therefore, the probability of drawing two *or fewer* heads in a flip of ten independent trials is approximately  $(4.4\% + 1.0\% + 0.1\%)$ , or 5.5%.<sup>69</sup> In employment discrimination cases, this would be an unsuccessful proof of discrimination since courts have interpreted *Hazelwood* and *Castaneda* to require at most a 5% chance of random sampling.<sup>70</sup>

65. *Id.*

66. The only possibilities are:

HHHTTTTTT,	HTHTTTTTT,	HHTHTTTTT,	HTTHTTTTT,	HTTTHTTTTT,
HHTTTTHTTT,	HTTTTTHTTT,	HTTTTTTHT,	HTTTTTTTH,	THHTTTTTTT,
THHTTTTTTT,	THHTTTTTTT,	THTTTHTTTT,	THTTTHTTTT,	THTTTTHHTT,
THTTTTHTTT,	THTTTTTTTH,	TTHHTTTTTT,	TTHHTTTTTT,	TTHHTTTTTT,
TTHTTTHTTT,	TTHTTTTHTT,	TTHTTTTTHT,	TTHTTTTTTH,	TTTHHTTTTT,
TTTHTHTTTT,	TTTHTHTTTT,	TTTHTTTTHT,	TTTHTTTTHT,	TTTHTTTTTH,
TTTTHHTTTT,	TTTTHHTTTT,	TTTTHTTTHT,	TTTTHTTTHT,	TTTTHTTTTH,
TTTTHTHTTT,	TTTTHTHTTT,	TTTTHTTHTT,	TTTTHTTTTH,	TTTTTHHTTT,
TTTTTHTHTT,	TTTTTHTTTH,	TTTTTTHTHT,	TTTTTTHTTH,	TTTTTTHTTH,

— for a total of forty-five possibilities.

67.  $(10!)/((10 - 1)! \cdot 1!) \cdot (1/2)^{10} = 10/1024$ .

68. The binomial distribution is not even needed for this calculation, although it obviously can be done. Using the probability rule, the probability of getting ten straight tails with no heads is simply  $1/2$  to the tenth power, or  $1/1024$ . See HAUSNER, *supra* note 38, at 97.

69. See CHARLES M. GRINSTEAD & J. LAURIE SNELL, INTRODUCTION TO PROBABILITY 193-95 (1997) (describing the summation rule). The key difference to note here is that the summation rule centers around the word "or" while the product rule centers around the word "and." Therefore, the probability that A or B or C happens is simply  $(P(A) + P(B) + P(C))$ , but the probability of A and B and C happening is  $(P(A) \cdot P(B) \cdot P(C))$ . See *id.*

70. Both *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 311 n.17 (1977), and *Castaneda v. Partida*, 430 U.S. 482, 497 n.17 (1976), state that discrimination can be assumed at "2 or 3 standard deviations." At two deviations, there is a 5% chance that the disparity is caused by randomness. Most

A problem arises, however, when the number of “successes” is fairly large because the binomial computations become exceedingly tedious.<sup>71</sup> For example, in *Castaneda*, since there were 339 jury members that were Mexican-American, it would have been somewhat difficult and time consuming<sup>72</sup> to use binomial distributions to exactly calculate the probability that 339 members were chosen, that 338 members were chosen, that 337 members were chosen, and so on.<sup>73</sup> Therefore, *Castaneda* implicitly used a “normal distribution” as an approximation of the binomial distribution.<sup>74</sup> According to the “central limit theorem,” normal distributions and their bell curves are an extremely accurate approximation of the binomial distribution as long as the number of trials ( $n$ ) multiplied by the probability of a success ( $p$ )<sup>75</sup> is equal to at least five.<sup>76</sup> It is important to note that while *Hazelwood* implicitly endorses the usage of these approximations, the number of trials and the probability of a minority hiring in that particular case do not justify using normal distributions as an approximation.<sup>77</sup> Similarly, many courts following *Hazelwood* have inappropriately used these approximations when the number of hiring decisions and/or the probability of a minority being hired were not high enough.<sup>78</sup>

Nevertheless, if it is appropriate to use normal distributions as approximations of binomial distributions, then the “standard deviation” of the normal distribution is used to determine the probability that a

---

courts have followed this 5% rule. See, e.g., Haggard, *supra* note 3, at 84.

71. Sugrue & Fairley, *supra* note 22, at 931; see also ROBERT L. WINKLER & WILLIAM L. HAYS, STATISTICS: PROBABILITY, INFERENCE, AND DECISION 226 (1975).

72. Though, in this Author’s view, not at all impossible. In fact, a computer program could be written to solve this problem with relative ease.

73. Sugrue & Fairley, *supra* note 22, at 931; see also WINKLER & HAYS, *supra* note 71.

74. Sugrue & Fairley, *supra* note 22, at 931–32; see also SAMUEL ESTREICHER & MICHAEL C. HARPER, CASES AND MATERIALS ON EMPLOYMENT DISCRIMINATION LAW 68 (2004). An entire discussion about normal distributions and why they are an accurate approximation of binomial distribution is beyond the scope of this Note. For more information about normal distributions, see HAUSNER, *supra* note 38, at 244.

75. Or the probability of a failure, given by  $(1 - p)$ , but in employment discrimination cases,  $p$  will usually be smaller than  $(1 - p)$  anyway.

76. This number five is not actually part of the central limit theorem, but has been accepted by most mathematicians as an appropriate value. See HAUSNER, *supra* note 38, at 244.

77. The percentage of the minorities in the labor market was either 15.4% or 5.7%. See *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 310–11 (1977). However, either of these values multiplied by the number of employees hired, or fifteen, is far less than five. See *id.* In this Author’s view, it is difficult to understand why *Hazelwood* did not simply use the binomial distribution, which did not have overly burdensome computations.

78. Sugrue & Fairley, *supra* note 22, at 926–27; see also, e.g., *Wilkins v. Univ. of Houston*, 654 F.2d 388 (5th Cir. 1981); *EEOC v. Am. Nat’l Bank*, 652 F.2d 1176 (4th Cir. 1981); *Hameed v. Int’l Ass’n of Bridge, Structural & Iron Workers Local 396*, 637 F.2d 506 (8th Cir. 1980); *Bd. of Educ. v. Califano*, 584 F.2d 576 (2d Cir. 1978); *Otero v. Mesa County Valley Sch. Dist.*, 568 F.2d 1312 (10th Cir. 1977); see also JOHN J. DONOHUE III, FOUNDATIONS OF EMPLOYMENT DISCRIMINATION LAW 297–98 (2003).

disparity in numbers is due to randomness.<sup>79</sup> The standard deviation of a normal distribution is described as the square root of  $((n \cdot p \cdot (1 - p)))$ , where  $n$  is the total number of people in the drawing pool,  $p$  is the probability of a success, and  $(1 - p)$  is the probability of a failure.<sup>80</sup> If the standard deviation is exactly one, then there is about a 32% chance that the selections were due to randomness; if the value is two, then the probability is 5% that it is due to randomness; and if the value is three, then the probability of randomness is 0.3%.<sup>81</sup>

### III. THE ERRONEOUS ASSUMPTION OF INDEPENDENT TRIALS MADE BY BINOMIAL DISTRIBUTIONS

As stated above, the binomial test relies on the “product rule.”<sup>82</sup> The product rule states that the probability of each of the events is independent of each other event<sup>83</sup> and the probability of a “success” is strictly based upon the population of the “successes” in relation to the total population set.<sup>84</sup> This implies that courts using binomial distributions make a critical assumption based on random sampling in that *every new hiring decision made by the employer is random and independent of every previous hiring decision*.<sup>85</sup> Unfortunately, this assumption is seriously flawed because “random sampling, which gives rise to the applicability of the product rule, is rarely present in nature or human affairs.”<sup>86</sup>

---

79. ROBERT BELTON ET AL., EMPLOYMENT DISCRIMINATION: CASES AND MATERIALS ON EQUALITY IN THE WORKPLACE 227 (2004) (“Generally, the fewer the number of standard deviations that separate an observed from a predicted result, the more likely it is that the observed disparity is not really a ‘disparity’ at all, but rather a random or chance fluctuation. Conversely, the greater the number of standard deviations, the less likely it is that chance is the cause of any difference between the expected and observed results.”); see also ESTREICHER & HARPER, *supra* note 74, at 68–69; HAGGARD, *supra* note 3, at 85; HAUSNER, *supra* note 38, at 210–12; PAETZOLD & WILLBORN, *supra* note 6, at 34; Meier et al., *supra* note 47, at 145.

80. ESTREICHER & HARPER, *supra* note 74, at 68–69; HAGGARD, *supra* note 3, at 85; HAUSNER, *supra* note 38, at 210–12; PAETZOLD & WILLBORN, *supra* note 6, at 34; Meier et al., *supra* note 47, at 144.

81. HAUSNER, *supra* note 38, at 210–12; PAETZOLD & WILLBORN, *supra* note 6, at 34–35; see also ESTREICHER & HARPER, *supra* note 74, at 68–69. As described *infra*, the Court in *Casteneda* would find an inference of discrimination somewhere between two and three standard deviations. However, as illustrated, there is a significant difference between two and three standard deviations, which begs the question of which value is appropriate. See generally Kaye, *supra* note 48.

82. See HAUSNER, *supra* note 38, at 97.

83. Normal distributions also require that the trials be independent. HAUSNER, *supra* note 38, at 244.

84. Recall that in the binomial distribution formula above, the total population set is illustrated by “ $n$ ,” and since there can only be two outcomes in binomial distribution analysis, the number of “failures” is  $(n - k)$ .

85. See HAUSNER, *supra* note 38, at 249–50; see also PAETZOLD & WILLBORN, *supra* note 6, at 32–33; Meier et al., *supra* note 47, 148–50.

86. Meier et al., *supra* note 47, at 153 (emphasis omitted).

#### A. HYPERGEOMETRIC DISTRIBUTIONS HELP SOLVE SOME INDEPENDENCE PROBLEMS

As explained above, binomial distributions require that each event is completely independent of every other event.<sup>87</sup> However, there is technically *never* complete independence within the employment discrimination context.<sup>88</sup> Recall that the binomial distribution is centered around the “product rule” of independent events, which states that the probability of  $n$  events occurring in a certain order is simply the multiplication of the probability of each individual event happening.<sup>89</sup> For example, the probability of rolling a one, then a two, then a three exactly in that order in three rolls on a six-sided dice is simply  $(1/6 \cdot 1/6 \cdot 1/6)$ , or  $1/216$ .

However, the same result is not reached in the employment discrimination context. Analogous to the previous example, imagine six employees numbered one through six applying for three distinct spots in a company.<sup>90</sup> The probability of the hiring of number one, then number two, then number three in exactly that order is  $(1/6 \cdot 1/5 \cdot 1/4)$ , or  $1/120$ , almost twice the probability of the previous example.<sup>91</sup> The reason for this is because once number one got hired, number two had to compete with one fewer persons for the second job, so the probability of number two getting hired given that number one was already hired increased from  $1/6$  to  $1/5$ .<sup>92</sup> Therefore, the hiring of number two is not independent of the hiring of number one. Hence, since the events are not independent of each other, the actual distribution that should be used in the employment context is “hypergeometric distribution,” or “sampling without replacement.”<sup>93</sup>

The hypergeometric distribution is a discrete probability distribution that describes the probability of a number of successes in a sequence of  $n$  draws from a finite population with replacement and is mathematically described as

---

87. HAGGARD, *supra* note 3, at 85; HAUSNER, *supra* note 38, at 210–12; PAETZOLD & WILLBORN, *supra* note 6, at 48; Meier et al., *supra* note 47, at 145.

88. PAETZOLD & WILLBORN, *supra* note 6, at 48–49.

89. See *supra* Part II.

90. For simplicity, I am assuming that order *does* matter. In the employment context, this means that the three positions open could mean that there are three distinct positions, such as jobs with different wages.

91. This is based on the product rule without independent trials. See HAUSNER, *supra* note 38, at 97.

92. Of course, the chance of number two getting hired in *any* position decreases once number one gets hired. The  $1/5$  is only the probability that number two gets hired for that particular second position, compared to only a  $1/6$  chance of a dice being rolled as a number two on the second particular roll.

93. GRINSTEAD & SNELL, *supra* note 69; see also PAETZOLD & WILLBORN, *supra* note 6, at 48–49.

$$f(k; N, D, n) = \frac{\binom{D}{k} \binom{N-D}{n-k}}{\binom{N}{n}}$$

where  $D$  is the total number of “successes” in the pool,  $N$  is the total number of objects being drawn from,  $n$  is the number of actual objects drawn, and  $k$  is the number of successes that are observed.<sup>94</sup> The reasoning behind the formula is that there are

$$\binom{N}{n}$$

—or  $(N!/(n! \cdot (N-n)!))$ —ways to select  $n$  objects from a whole set of  $N$  objects,<sup>95</sup> there are

$$\binom{D}{k}$$

—or  $(D!/(k! \cdot (D-k)!))$ <sup>96</sup>—ways to obtain  $k$  successes; and there are

$$\binom{N-D}{n-k}$$

—or  $(N-D)!/((n-k)! \cdot ((N-D)-(n-k))!)$ —ways to fill the rest of the sample with nondefective objects.<sup>97</sup> Understandably, like binomial distributions, this formula can be extremely confusing,<sup>98</sup> so an example involving systemic employment discrimination will hopefully make it a little more clear.

Suppose that an employer has four male and six female applicants for five identical openings.<sup>99</sup> Absent any discrimination or other reasoning for favoring one individual over any other individual, the

94. GRINSTEAD & SNELL, *supra* note 69.

95. *Id.*

96. *Id.* This is assuming that order does not matter in selections, or in the systemic employment context, that the open positions are no different from one another.

97. *Id.*

98. See *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 318 (1977) (Stevens, J., dissenting) (illustrating the difficulty for lawyers and judges to apply these mathematical formulas); DONOHUE, *supra* note 78, at 297–98.

99. It is essential that the openings be exactly identical, or else the hypergeometric (and for that matter, binomial) distributions would not work because then *order would matter*.

expected value of the number of female hirings is 60% of five, or three.<sup>100</sup> However, the employer decides to hire only two women, therefore filling the openings with only 40% women, despite the fact that 60% of the applicant pool was female. Hypergeometric distributions must be analyzed to determine if this discrepancy is due to random number samplings or if it is due to other factors such as discrimination.<sup>101</sup> First, hypergeometric functions should be used to determine the probability of *exactly* two women being hired. There are

$$\binom{D}{k}$$

—or  $(6!/(2! \cdot (6-2)!))$ —or fifteen ways to pick two women out of the six women in the applicant pool,<sup>102</sup> which can also be established in this simple example using common sense.<sup>103</sup> There are

$$\binom{N-D}{n-k}$$

—or  $((10-6)!/((5-2)! \cdot ((10-6)-(5-2))!))$ —or four ways of picking the remaining three positions between the four male candidates,<sup>104</sup> which can also be determined by listing the possibilities.<sup>105</sup> By the product rule, there are  $(15 \cdot 4)$  or sixty ways, to choose two women amongst a pool of five women and three men among a pool of five men for five openings.<sup>106</sup> The number of total ways to pick any five individuals amongst a total pool of ten is simply

$$\binom{N}{n}$$

—or  $(10!/(5! \cdot 5!))$ —or 252.<sup>107</sup> Therefore, the probability of picking

100. HAUSNER, *supra* note 38, at 168–70.

101. See PAETZOLD & WILLBORN, *supra* note 6, at 48–49.

102. GRINSTEAD & SNELL, *supra* note 69, at 93–96.

103. Let us suppose the five women are A, B, C, D, E, and F. The only possible combinations (where order is not taken into account) are AB, AC, AD, AE, AF, BC, BD, BE, BF, CD, CE, CF, DE, DF, and EF. Note that this counting method can only be done since the numbers chosen are so small. Note also that we have to individualize each woman since we are dealing with sampling with replacement, and therefore each sample is not independent of each other.

104. GRINSTEAD & SNELL, *supra* note 69, at 93–96.

105. Let us suppose the four male candidates are A, B, C, and D. The only possible combinations (where order is not taken into account) are ABC, ABD, ACD, and BCD. Note that this counting method can only be done since the numbers chosen are so small. Note also that we have to individualize each male since we are dealing with sampling with replacement, and therefore each sample is not independent of each other.

106. GRINSTEAD & SNELL, *supra* note 69, at 31.

107. Note that this is the binomial coefficient fully described in Part II of this Note. As described



*exactly* two women for this particular position would be 60/252 (approximately 23.8%).

However, this percentage understates the probability in the discrimination context because the appropriate question is not the probability of hiring *exactly* two women, but rather the probability that the employer will employ two *or fewer* women.<sup>108</sup> Therefore, hypergeometric functions must also be applied to the probability that the employer hires only one woman, in this case approximately 2.4%,<sup>109</sup> and the probability that the employer hires no women at all, which is 0%.<sup>110</sup> Thus, the probability of an employer hiring two or fewer women in our example is (23.8% + 2.4% + 0%),<sup>111</sup> or approximately 26.2%. Although this implies that the discrepancy between the employer's percentage of women employees and the percentage of women in the applicant pool has a decent probability of being due to chance,<sup>112</sup> a different outcome is conceivable when erroneously using binomial distributions, as advocated by *Castaneda*<sup>113</sup> and *Hazelwood*.<sup>114</sup>

Recall the formula for binomial distributions as

$$f(k; n, p) = \binom{n}{k} p^k (1 - p)^{(n-k)}$$

where  $p$  is the probability of the hiring of a woman,  $(1 - p)$  is the probability of the hiring of a man,  $n$  is the total number of applicants being hired, and  $k$  is the number of women who were actually hired.<sup>115</sup> The probability of a woman being hired ( $p$ ) is proportional to the number of women in the applicant pool,<sup>116</sup> or in our example 60%. Therefore, the probability that a woman is not hired  $(1 - p)$  is 40%. Therefore, the binomial distribution implies that the probability that the first two hires are women is  $(3/5)^{(2)}$ , or 9/25, and the probability that the

there using the "BEN" example, this is assuming that order does not matter. For example, if the five open positions were for five distinct job titles, then the total number would instead be  $(10!/5!)$  instead of  $(10!/(5! \cdot 5!))$ .

108. See Sugrue & Fairley, *supra* note 22, at 931.

109.  $(6!/(1! \cdot (6 - 1)!)) \cdot (10 - 6)!/((5 - 1)! \cdot ((10 - 6) - (5 - 1))!)/(10!/(5! \cdot 5!)) = 6/252$ .

110. Although the hypergeometric function could be used to get this amount, simple common sense shows that it is impossible to hire all men for five positions when there are only four men applying.

111. See *supra* note 67.

112. Or at a minimum, a plaintiff would not be able to use these statistics of proof of gender discrimination. See *supra* note 66.

113. *Castaneda v. Partida*, 430 U.S. 482, 497 n.17 (1976).

114. *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 312 n.17 (1977).

115. *Infra* Part II.

116. This is based on expected value. See *Hazelwood*, 433 U.S. at 312 n.17. Note that this is because the binomial distribution assumes there are no other factors taken into account. See GRINSTEAD & SNELL, *supra* note 69, at 98-101.

last three hires are men is  $(2/5)^3$ , or  $8/125$ .<sup>117</sup> Once again following the product rule, the binomial theorem therefore provides that the probability of women for the first two hires and men for the last three hires is  $(9/25 \cdot 8/125)$ , or approximately 2.3%. However, the hiring does not have to occur in this particular order, and there are

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

—or  $(5!/(2! \cdot 3!)) = 10$ —ways of picking two women and three men from an applicant pool of ten, which can also be shown by simply listing the possibilities.<sup>118</sup> Therefore, the binomial distribution implies that there is a  $(10 \cdot 2.3\%)$ , or approximately 23.0%, chance that the employer hires *exactly* two women, which is different from the true 23.8% value determined by hypergeometric functions.

Completing the example, the probabilities of selecting exactly one woman and exactly no women are also analyzed using the binomial theorem to find the probability that the employer would hire two *or fewer* women. The probability that the employer employs only one woman is approximately 7.7%<sup>119</sup> and the probability that the employer hires no women is 1.0%.<sup>120</sup> Therefore, the binomial theorem produces a result of  $(23.0\% + 7.7\% + 1.0\%)$ , or approximately 31.7%.<sup>121</sup> Although either method yields a result in this particular situation that would draw an inference of no discrimination, the main point is that the difference between the two distributions, in this case 31.7% and 26.2%, can lead to extremely significant results where “the binomial model errors on the ‘conservative’ side from a plaintiff’s point of view.”<sup>122</sup> In other words, this erroneous discrepancy will side against the plaintiff in overstating the probability of chance.<sup>123</sup> This corresponding understatement of discrimination and overstatement of random chance is why “[i]t is critical that the process be modelled [sic] correctly so that appropriate inferences

117. This is incorrect because the sample sets are being erroneously replaced.

118. Let’s call the women W and the men M. The only possibilities are WWMMM, WMWMM, WMMWM, WMMM, MWMM, MWMW, MMWW, MMWM, and MMMW. Note that it is not necessary to individualize the men and women since the binomial theorem assumes that the selection of an individual is independent of the selection of any other individual. See HAUSNER, *supra* note 38, at 252–53.

119.  $(3/5)^1 \cdot (2/5)^4 \cdot (5!/(1! \cdot 4!)) = 7.68\%$ .

120. Although this can be found using binomial distributions, this can also be found using the simple product rule of individual trials. GRINSTEAD & SNELL, *supra* note 69, at 35–37. Therefore, this amount would be  $(2/5)^5$  or 1.024%. This should be an obvious error to the reader. It is impossible to hire four men for five positions, however, the binomial theorem lets this be done because it assumes that every event is independent of the other.

121. See *supra* note 107.

122. Sugrue & Fairley, *supra* note 22, at 926–27.

123. See *id.*

can be made."<sup>124</sup>

The only reason why *Castaneda*, *Teamsters*, and *Hazelwood* were able to use binomial distributions and the standard deviations based on the accurate approximation of binomial distributions through normal distributions<sup>125</sup> is that the drawing pool of minorities in each of those cases was extremely large, indicating that the binomial distribution was a decent approximation of the hypergeometric distribution.<sup>126</sup> Having a large number of candidates to choose from, combined with a comparatively small number of hiring decisions, creates a binomial estimate very close to the true hypergeometric distribution. Even though the employment decisions remain somewhat dependent on each other, because there is no sample replacing, the decisions are *less* dependent on each other.<sup>127</sup>

Using a binomial distribution as an accurate approximation of hypergeometric distribution in cases with high selection pools and relatively low hiring decisions can best be illustrated using the "dice" example above. Since employment decisions are not independent of one another, the probability of number two getting hired out of six candidates for the second position given that number one has already been hired is 1/5 compared to 1/6 which the binomial theorem requires; a very considerable difference.<sup>128</sup> However, if there were sixty-thousand candidates instead of only six, the probability of number one being hired for the first position would be 1/60,000 and the probability of number two being hired for the second position given that number one has already been hired is now 1/59,999. These two probabilities are so close that the dependence between the hiring decisions of number one and number two can nearly be considered independent, therefore moving the distribution from a hypergeometric to a binomial distribution.<sup>129</sup> For example, *Castaneda* compared the population of 870 summoned grand jurors over an eleven year period to the entire population of Hidalgo County at over 180,000;<sup>130</sup> *Teamsters* used entire major cities as the appropriate bases of comparison to only a few thousand employees;<sup>131</sup> and *Hazelwood* utilized percentages based upon over 19,000 teachers in the local labor market compared to slightly over four hundred hired by

---

124. PAETZOLD & WILLBORN, *supra* note 6, at 49 n.1.

125. Recall that normal distributions can be used as accurate approximations of binomial distributions based on the central limit theorem under certain circumstances. *See supra* Part II.

126. *See* PAETZOLD & WILLBORN, *supra* note 6, at 49; Sugrue & Fairley, *supra* note 22, at 938.

127. *See* Sugrue & Fairley, *supra* note 22, at 938.

128. The difference between 1/5 and 1/6 is over 3%. *See* Sugrue & Fairley, *supra* note 22, at 938 ("[I]f blacks, for example, are selected at a disproportionately low rate, the percentage of blacks remaining in the eligible pool will tend to increase as selections are made.").

129. *See* PAETZOLD & WILLBORN, *supra* note 6, at 49; Sugrue & Fairley, *supra* note 22, at 938.

130. *Castaneda v. Partida*, 430 U.S. 482, 496-97 n.17 (1977).

131. *Int'l Bhd. of Teamsters v. United States*, 431 U.S. 324, 337-38 n.17 (1977).

that particular employer.<sup>132</sup>

None of these cases illustrates the important fact that binomial distributions, and the approximations through normal distributions and the resulting standard deviations, could greatly differ from the accurate hypergeometric distributions.<sup>133</sup> Unfortunately, many lower courts have blindly followed the formulas put forth in these cases without truly analyzing whether they actually fit.<sup>134</sup> The damage caused by this in jury selection cases such as *Casteneda* is not as significant because most jury selection pools are large. The reasons they are so large are that most adults are eligible for jury selection, the distances between the court houses and residences are not as significant, and the number of people chosen to be on a jury is relatively small.<sup>135</sup> On the other hand, choosing to use binomial analysis instead of hypergeometric statistics in systemic employment discrimination cases where the number of hiring decisions is a substantial percentage of the eligible pool could produce disastrous results. This, in turn, could “overstate the likelihood that differences in selection rates could be attributed to chance and understate the statistical significance of the racial disparities observed.”<sup>136</sup> Therefore, using binomial distributions inappropriately as approximations for hypergeometric distributions would cause an understatement of discrimination, making these figures extremely unreliable.<sup>137</sup>

B. THERE ARE OTHER INDEPENDENCE PROBLEMS THAT ARISE REGARDLESS OF WHETHER BINOMIAL OR HYPERGEOMETRIC DISTRIBUTIONS ARE USED

Unfortunately, hypergeometric distributions do not solve all problems stemming from a lack of independence. Unlike instances of jury selection, another common reason why each employment decision is not random is that “[n]o employer truly hires at random from a proxy pool.”<sup>138</sup> The two most common “pooling” problems are “grouping” problems and differences in preference between classes.

1. *Grouping Effects Lead to Nonindependent Employment Decisions*

There are numerous “grouping” scenarios where the group that the employer selects from is not completely random, such as when “a person

---

132. *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 303 (1977). The Court in *Hazelwood* also rejected the Eighth Circuit’s argument that the entire workforce of over one thousand for the employers should be used for unrelated reasons. *Id.* at 310–11.

133. See PAETZOLD & WILLBORN, *supra* note 6, at 46–49; Sugrue & Fairley, *supra* note 22, at 938.

134. See sources cited *supra* note 19.

135. See Sugrue & Fairley, *supra* note 22, at 938.

136. *Id.*; see also PLAYER, *supra* note 10, at 352 (“A major factor influencing the probative value of a statistical analysis is the size of the statistical sample. Generally, the smaller the sample size the less reliable the inference of discrimination.”).

137. PAETZOLD & WILLBORN, *supra* note 6, at 49; Sugrue & Fairley, *supra* note 22, at 938.

138. PAETZOLD & WILLBORN, *supra* note 6, at 50.

recruited . . . recommend[s] or otherwise bring[s] along friends into the company [or] success with applicants from a given school or other organization may set up a short-term 'pipeline' of future applicants."<sup>139</sup> These grouping scenarios "might be entirely compatible with nondiscriminatory hiring" and is oftentimes the "rule in real world hiring, not the exception."<sup>140</sup> This grouping effect makes the next employment decision likely to be skewed from the expected value of the disadvantaged group.<sup>141</sup>

A good illustration of the grouping effect can be shown in an extreme case, *EEOC v. Consolidated Service Systems*.<sup>142</sup> In *Consolidated Service Systems*, the employer of about one hundred cleaners had an 81% workforce consisting of Korean-Americans, despite the fact that Korean-Americans only made up 1% of the general population and 3% of the workforce engaged in this type of business in the metropolitan area where the service operated.<sup>143</sup> The statistical analysis methods advocated by *Teamsters* and *Hazelwood* would definitely lead to an inference of discrimination.<sup>144</sup> Fortunately, the Court recognized that there was not independence in the hiring decisions since the employer's hiring offices were in a heavily Korean-American populated neighborhood and that most of the applicants came through word-of-mouth recruiting.<sup>145</sup> Therefore, when one Korean-American was hired, it was more likely that another Korean-American would be hired in the next application decision due to this reference-style recruiting.<sup>146</sup>

In *Consolidated Service Systems*, the case was such an extreme example of grouping that the court was able to recognize the nondiscriminatory reason for the disparity in the statistics. However, some form of grouping almost always occurs in employment decisions, and it usually is not detected.<sup>147</sup> The lack of independence between hiring decisions due to grouping can skew statistics toward either the plaintiff or the defendant.<sup>148</sup> Unfortunately, many courts do not recognize the significant effects that grouping can have on statistics. Furthermore,

---

139. Meier et al., *supra* note 47, at 154; see also ESTREICHER & HARPER, *supra* note 74, at 78.

140. Meier et al., *supra* note 47, at 154.

141. *Id.*; see also PAETZOLD & WILLBORN, *supra* note 6, at 49-50.

142. 989 F.2d 233 (7th Cir. 1993); see also ESTREICHER & HARPER, *supra* note 74, at 78.

143. *Consolidated Serv. Sys.*, 989 F.2d at 234-35.

144. Unfortunately, the opinion in *Consolidated Service Systems* does not state exactly how many employees there were, only the percentages. Therefore, an exact binomial distribution or standard deviation cannot be obtained. However, based on the percentages, it is extremely likely that the binomial distributions or standard deviations would have shown that the chance that these hiring decisions were due to chance would be less than 5%. *Id.*; see also GRINSTEAD & SNELL, *supra* note 69, at 97-99 (describing how to calculate binomial probabilities).

145. *Consolidated Serv. Sys.*, 989 F.2d at 236-38.

146. *Id.*; see also Meier et al., *supra* note 47, at 143.

147. See Meier et al., *supra* note 47, at 143.

148. *Id.*

these grouping problems would arise regardless of whether binomial or hypergeometric distributions are used.

## 2. Protected Class Preference

Many times statistically significant disparities in race or sex are explained by disparate interest in a particular field or position rather than by actual discrimination.<sup>149</sup> Therefore, the applicant pool is not representative of the percentages of classes in the general population.<sup>150</sup> This problem is especially apparent when the general population is used as the “pool” instead of just the pool of applicants, such as in *Teamsters*. The cases that use a comparison of the workforce disparities to the general population many times ignore the simple fact that the underrepresented classes simply do not have as much of an interest in the job or location of the employment opportunity. Once again, using hypergeometric distributions would not help solve this problem of underrepresented class preference.

## C. MORE WEIGHT SHOULD BE PUT ON NONSTATISTICAL ANALYSIS

Statistical evidence is the primary, and oftentimes exclusive, evidence in systemic employment discrimination cases because it is commonly viewed as the easiest, most concrete, and most thorough method of proving a *prima facie* case of discrimination.<sup>151</sup> Similarly, courts have found that individual and isolated evidence of discrimination should not be used with much weight in establishing an inference of discrimination in systemic treatment cases since “individual acts of discrimination are not enough to meet this standard.”<sup>152</sup>

I believe the opposite logic should be exercised; instead of statistical evidence providing the primary support with individual treatment as secondary evidence, specific cases and instances should be used as central evidence with statistical evidence only persuasive in showing discrimination.<sup>153</sup> Specific and concrete evidence of discrimination should

---

149. See *EEOC v. Sears, Roebuck, & Co.*, 839 F.2d 302, 324 (7th Cir. 1988); BELTON, *supra* note 7, at 229; ZIMMER ET AL., *supra* note 31, at 279; Kingsley R. Browne, *Statistical Proof of Discrimination: Beyond Damned Lies*, 68 WASH. L. REV. 477, 510–12 (1993).

150. ZIMMER ET AL., *supra* note 31, at 267 (“General population data can be deceptive. Suppose an employer’s typical pool is 99 percent female or a mining company’s complement of miners is 99 percent male. Is it appropriate to draw an inference of intent to discriminate in these cases?”); see also *Int’l Bhd. of Teamsters v. United States*, 431 U.S. 324, 399–400 n.20 (1977); DAVID C. BALDUS & JAMES W.L. COLE, *STATISTICAL PROOF OF DISCRIMINATION* 4.1–.3 (1980); BELTON, *supra* note 7, at 228; Judith P. Vladeck, *Sex Discrimination in Higher Education*, 4 WOMEN’S RTS. L. REP. 59, 74–75 (1978).

151. Meier et al., *supra* note 47, at 139–40; see also, e.g., *EEOC v. Chicago Miniature Lamp Works*, 947 F.2d 292, 302 (7th Cir. 1991); *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 311 n.17 (1977); *Teamsters*, 431 U.S. at 335–38; PAETZOLD & WILLBORN, *supra* note 6.

152. PAETZOLD & WILLBORN, *supra* note 6.

153. There exists the belief that statistical data should hold even less weight in employment discrimination cases, and should instead act as “merely an adjunct to evidence.” Browne, *supra* note 149, at 477.

not be used to “simply bolster the statistics.”<sup>154</sup> This Note has illustrated that only a limited amount of faith should be instilled in using statistics and, therefore, should not be used as a primary tool for proving discrimination. However, there must be some way of proving a *prima facie* case of discrimination, and to do this, more weight should be put on individual evidence and occurrences. If the employer discriminated against one individual in a disadvantaged minority group, the employer will likely discriminate against others in the group. In these cases, statistical evidence should only be supportive and persuasive of the individual evidence.

### CONCLUSION

*Teamsters* and *Hazelwood* used statistical data from standard deviations based on binomial distributions to prove an inference of employment discrimination in systemic disparate treatment cases. Lower cases have since used these standard deviations to create inferences of discrimination.<sup>155</sup> However, “lawyers and judges are often asked to resolve difficult quantitative issues that lie far from their area of expertise,” and therefore are not aware of, or improperly ignore, important assumptions made by these statistical methods.<sup>156</sup> Binomial distributions assume that each employment decision is independent of every other decision. Because employment decisions involve “sampling without replacement,” these decisions are therefore not independent, and a hypergeometric model instead of binomial distributions should be used. However, even if hypergeometric distributions are used or if the binomial distribution provides an accurate approximation of the hypergeometric distribution, the employment decisions are still not random due to grouping effects or minority preference which skew the randomness of the pool. Therefore, other evidence, such as individual instances of discrimination, should replace statistical data as the primary source of evidence in showing discrimination in systemic disparate treatment cases.

---

154. ROTHSTEIN & LIEBMAN, *supra* note 7.

155. See sources cited *supra* note 19.

156. DONOHUE, *supra* note 78.